Journal of Agricultural Economics and Development

Homepage: https://jead.um.ac.ir

Research Article
Vol. 37, No. 3, Fall 2023, p. 289-307

Applying Hybrid Models Based on Deep Machine Learning in Smart Agriculture (Case study: Forecasting the Future Price of Pistachios)

M.R. Haj-Seyedjavadi [1], R. Heydari [2*]

Received: 11-11-2022
Revised: 13-03-2023
Accepted: 10-04-2023
Available Online: 02-05-2023

How to cite this article:
Haj-Seyedjavadi, M.R., & Heydari, R. (2023). Applying hybrid models based on deep machine learning in smart agriculture (Case study: forecasting the future price of pistachios). *Journal of Agricultural Economics & Development, 37*(3), 289-307. (In Persian with English abstract). https://doi.org/10.22067/jead.2023.79545.1165

Introduction

The agricultural sector is one of the most basic and vital component in the social and economic structures of any country. Today, with increasing in the world's population and needing to provide food on the other hand, and increasing in the price fluctuations of agricultural products on the other hand, traditional agriculture is no longer responsible for the sustainable food security of the world population. In recent years, the occurrence of two incidents of the spread of the corona virus and the outbreak of war in Ukraine, have made the price of agricultural products extremely unstable. Today, even many farmers and agricultural associations in developing countries are not aware of the changes in market prices and the latest technological developments in the field of agricultural product prices, and they do not have the ability to discover the optimal price for selling their products. In such a situation, the use of intelligent models in order to accurately forecast the price of agricultural goods is vitally important for farmers and agricultural sector activists.

Smart agriculture is an emerging concept that involves the integration of advanced technologies to collect and analyze data in order to solve the challenges and problems of the agricultural sector. In the meantime, forecasting the price of agricultural products involves with some basic challenges; including: 1) Data of agricultural product price is mostly non-linear, unstable, non-normal, and noisy and follows chaotic behavior, 2) There is uncertainty in the forecasted data obtained from different models, 3) In the studies related to price forecasting, the "publishable base model" is not provided in order to provide the forecasted price values. Therefore, the aim of this study is to provide a non-linear hybrid intelligent model for accurate forecasting of the future price of pistachios in the field of smart agriculture through managing the multidimensional nature of data, considering uncertainty in the forecasting data and finally building a publishable base model in the field of product price prediction.

The hybrid model proposed in this study has the following innovations; 1) the deep learning neural network model and the Auto-Encoder network have been used to forecast the agricultural product price and determine the optimal lag of price as an input variable simultaneously, 2) The Monte Carlo method has been used as a non-parametric method to provide a confidence interval and calculate the most likely price that can happen, 3) The practical application of price forecasting models, i.e., "publishable base model" is presented in order to provide forecasted price values.

1- Researcher, Agricultural Planning, Economics and Rural Development Research Institute (APERDRI), Tehran, Iran
2- Assistant Professor, Agricultural Planning, Economics and Rural Development Research Institute (APERDRI), Tehran, Iran
(*- Corresponding Author Email: rezaheidari3631@gmail.com)
https://doi.org/10.22067/jead.2023.79545.1165

Materials and Methods

The implementation of the proposed hybrid model in this study includes the steps of "data preparation", "data feature engineering", "training and testing the final deep learning neural network model", "building the optimal base model", "creating the most likely price scenarios" using the Monte Carlo method and "inferring new prices or making out-of-sample forecasting" with new data sets" by feeding new price data into the deep learning neural network model. In the proposed hybrid model, data mining techniques are used, including Wavelet Transform (WT), Long-Short Term Memory (LSTM), Auto-Encoder network (AE), Monte Carlo-Markov chain (MCMC) simulation method and the concept of "inferring new prices".

In the data preparation stage, using methods such as data smoothing, data rebuilding, correction of duplicate data in several consecutive days, and correction of missing data, the continuous set of pistachio future price time series is prepared to enter the primary model. Also, the wavelet transform function has been used for de-noising the data, the Auto-Encoder network has been used to determine the optimal lag, the Monte Carlo-Markov chain simulation has been used to create the most probable price scenarios, and the deployment concept has been used for out-of-sample forecasting with new data sets. The data used in this study is the time series of the daily price of pistachio futures on the Iran Commodity Exchange in the period from 10/13/2019 to 12/14/2021 in Rials per kilogram.

Results and Discussion

The results of this study showed that 1) by using the wavelet theory to de-noise the data, the error rate of the price data was reduced and the data had a stable trend, 2) the results of the implementation of the Auto-Encoder network showed that the optimal lag of one can be used as an input variable to forecast the future price of pistachios, 3) The outcomes derived from employing Monte Carlo-Markov chain simulation, coupled with out-of-sample forecasting using the new dataset, reveal compelling insights into the future pricing of pistachios on the Iranian Commodity Exchange. According to the analysis, the most probable and sanguine projection places the future price at the price ceiling of 213 thousand Tomans. Impressively, the forecasted price exhibits a minimal variance of merely 0.7% from the actual observed price, attesting to the precision of the proposed model. The overall accuracy of the model stands commendably high at approximately 93%.

Conclusion

Based on the results, firstly, the forecasted price has a small error with the actual price and this small error shows the power of the built model in forecasting the future price trend of pistachios. Secondly, the alignment of the price resulting from the Monte Carlo simulation with the new price can also be used as a confidence index in risk management for traders and market participants. Thirdly, the process set is the most complete value chain in the production of price forecasting models. Therefore, the use of the proposed hybrid model and the use of the components used in it, i.e. wavelet transform function, Auto-Encoder network, deep learning neural network, Monte Carlo simulation and the concept of inferring new prices; are suggested.

Keywords: Datamining, Deep learning neural network, Future price forecast, Monte Carlo-Markov chain, Wavelet theory

Journal of Agricultural Economics and Development
https://jead.um.ac.ir

مقاله پژوهشی
جلد ۳۷، شماره ۳، پاییز ۱٤۰۲، ص ۳۰۷-۲۸۹

به کارگیری مدل‌های هیبریدی مبتنی بر یادگیری عمیق ماشین در کشاورزی هوشمند (مطالعه موردی: پیش‌بینی قیمت آتی پسته)*

سید محمدرضا حاج سیدجوادی[1]- رضا حیدری[2]

تاریخ دریافت: ۱۴۰۱/۰۸/۲۰
تاریخ بازنگری: ۱۴۰۱/۱۲/۲۲
تاریخ پذیرش: ۱۴۰۲/۰۱/۲۱

چکیده

امروزه بسیاری از کشاورزان و فعالان بخش کشاورزی از تغییرات قیمت‌های بازار و آخرین پیشرفت‌های فناوری در حوزه قیمت محصولات کشاورزی آگاهی‌های لازم را ندارند؛ بنابراین بهره‌گیری از مدل‌های هوشمند برای پیش‌بینی دقیق قیمت کالاهای کشاورزی در حوزه کشاورزی هوشمند برای آنها اهمیت حیاتی دارد. لذا هدف از این مطالعه، ارائه یک مدل هوشمند بر پایه داده‌کاوی از نوع هیبریدی غیرخطی برای پیش‌بینی دقیق قیمت آتی پسته به منظور رفع محدودیت‌های موجود شامل ماهیت چندبعدی داده‌ها، عدم قطعیت در داده‌های پیش‌بینی شده و نهایتاً ارائه و ساخت مدل پایه قابل انتشار در زمینه به کارگیری الگوریتم‌های یادگیری عمیق ماشین برای پیش‌بینی قیمت محصولات کشاورزی است. داده‌های مورد استفاده در این مطالعه سری زمانی قیمت روزانه سررسید قراردادهای آتی پسته فندقی در بورس کالای ایران در دوره زمانی ۱۳۹۸/۰۷/۲۱ تا ۱۴۰۰/۰۹/۲۳ است. نتایج حاصل از این مطالعه نشان داد که ۱) با بکارگیری تئوری موجک برای نوفه‌زدایی داده‌ها، میزان خطای داده‌های قیمت کاهش یافته و داده‌ها از یک روند باثبات برخوردار شدند، ۲) نتایج حاصل از اجرای شبکه کدکننده خودکار منتج به انتخاب وقفه بهینه یک، به‌عنوان متغیر ورودی برای پیش‌بینی قیمت آتی پسته تشخیص داده شد، ۳) نتایج حاصل از بکارگیری شبیه‌سازی مونت کارلو-زنجیره مارکف و نیز پیش‌بینی خارج از نمونه با مجموعه داده‌های جدید، بیانگر این است که محتمل‌ترین و خوشبینانه‌ترین قیمت قابل وقوع برای قیمت آتی پسته فندقی در بورس کالای ایران، در سقف قیمتی ۲۱۳ هزار تومان قرار دارد و قیمت پیش‌بینی شده با قیمت واقعی دارای اختلاف اندکی است (میزان خطا ۰/۷ درصد است). بر اساس نتایج حاصل شده، استفاده از مدل هیبریدی پیشنهاد شده و بکارگیری اجزای بکار برده شده در آن یعنی تابع تبدیل موجک، شبکه کدکننده خودکار، شبکه عصبی یادگیری عمیق، شبیه‌سازی مونت کارلو و استنتاج قیمت‌های جدید به‌عنوان کامل‌ترین زنجیره ارزش دو بخشی تحت یک مدل مرجع و پایه قابل انتشار برای پیش‌بینی و آزمون سایر محصولات کشاورزی با امکان به کارگیری تواترهای زمانی مختلف پیشنهاد می‌شود.

واژه‌های کلیدی: پیش‌بینی قیمت آتی پسته، تئوری موجک و شبیه‌سازی مونت کارلو-زنجیره مارکف، داده‌کاوی، یادگیری شبکه عصبی عمیق

۱- پژوهشگر، مؤسسه پژوهش‌های برنامه‌ریزی، اقتصاد کشاورزی و توسعه روستایی، تهران، ایران
۲- استادیار پژوهشی، موسسه پژوهش‌های برنامه‌ریزی، اقتصاد کشاورزی و توسعه روستایی، تهران، ایران
(*-نویسنده مسئول: Email: rezaheidari3631@gmail.com)

https://doi.org/10.22067/jead.2023.79545.1165

*- مطالعه حاضر برگرفته از طرح پژوهشی با عنوان «طراحی مدل پیش‌بینی قیمت آتی پسته در بورس کالای کشاورزی» در موسسه پژوهش‌های برنامه‌ریزی، اقتصاد کشاورزی و توسعه روستایی می‌باشد.

مقدمه

امروزه با افزایش جمعیت جهان و نیاز به تأمین غذا از یک طرف و افزایش قیمت نهاده‌های تولید و به تبع آن نوسانات شدید قیمت بازار محصولات کشاورزی، کشاورزی سنتی دیگر پاسخگوی امنیت پایدار غذایی جمعیت جهان نیست. تأمین غذای کافی و سالم برای جمعیت رو به رشد جهان از یک طرف و بی‌ثباتی بازار محصولات اساسی کشاورزی از طرف دیگر، همواره چالشی اساسی برای دولت‌ها، سیاست‌گذاران و اقتصاددانان بخش کشاوزی بوده است (Liu *et al.*, 2020). در این میان، قیمت محصولات کشاورزی به عنوان یکی از مولفه‌های مهم بخش کشاورزی، همواره ماهیتی ناپایدار داشته و توانسته است با افزایش یا کاهش خود بر اقتصاد کشاورزی تأثیرگذار باشد (Vohra *et al.*, 2019). طی سال‌های اخیر نیز وقوع دو رخداد انتشار ویروس کرونا و وقوع جنگ اوکراین، قیمت کالاهای کشاورزی را شدیداً ناپایدار کرده است (Heydari and Haj Seyed Javadi, 2022).

فارغ از بحران‌های جهانی موثر بر قیمت محصولات کشاورزی، امروزه حتی بسیاری از کشاورزان و انجمن‌های کشاورزی در کشورهای در حال توسعه از تغییرات قیمت‌های بازار و آخرین پیشرفت‌های فناوری در حوزه قیمت محصولات کشاورزی آگاهی‌های لازم را نداشته و توانایی کشف قیمت مطلوب برای فروش محصولات خود را ندارند (Joshi and Patel, 2022; Hegde *et al.*, 2021). بنابراین، ارائه آخرین دستاوردهای پیش‌بینی قیمت محصولات کشاورزی به شرکت‌ها، کشاورزان و کسب‌وکارهای کشاورزی کمک می‌نماید تا به اهداف قابل قبول خود دست یافته و با افزایش آگاهی آنان می‌توان تا حدودی ریسک قیمتی را کاهش داد. جدای از ضروت و اهمیت پیش‌بینی قیمت محصولات کشاورزی، ابزارها و تکنیک‌های مورد نیاز برای اجرای یک پیش‌بینی دقیق خود به یک مسئله و چالش واقعی برای کشاورزان تبدیل شده است (Joshi and Patel, 2022; Paul and Garai, 2021; Mitra and Paul, 2020; Vohra *et al.*, 2019).

کشاورزی هوشمند، مفهومی نوظهور است که شامل تلفیق فن‌آوری‌های پیشرفته مانند اینترنت اشیا، هوش مصنوعی و الگوریتم‌های یادگیری برای جمع‌آوری داده‌ها برای تجزیه و تحلیل بیشتر به فعالان بخش کشاورزی مطرح می‌گردد، تا اطلاعات دقیق را برای تصمیم‌گیری بهتر جهت کسب نتایج دقیق از محصول ارائه دهد. در این میان، تکنیک‌های یادگیری ماشین و الگوریتم‌های داده‌کاوی می‌تواند به‌عنوان یکی از گزینه‌های مطلوب در مسیر حل مسائل و مشکلات کشاورزی هوشمند (از جمله ارائه پیش‌بینی دقیق قیمت محصولات کشاورزی) مطرح شود؛ چرا که در دنیای امروز، کشاورزی با تولید حجم زیادی از داده‌ها روبرو است و به کمک محاسبات نرم می‌تواند با کشف الگوهای پنهان در ماهیت چندبعدی داده‌های سری زمانی کشاورزی به پیش‌بینی دقیق قیمت محصولات کشاورزی کمک نماید (Hirapara and Vanjara, 2022; Vohra *et al.*, 2019).

عموماً مجموعه داده‌های سری زمانی کشاورزی از جمله قیمت، ممکن است ماهیت چندبعدی داشته باشند. به بیان دیگر، عمدتاً داده‌های سری زمانی کشاورزی ماهیت غیرخطی، ناایستا، غیر نرمال، نویزدار داشته و از رفتار آشوبناک تبعیت می‌کند. اگرچه مدل‌های سنتی رایج محبوبیت زیادی را بواسطه آسانی در محاسبات کسب نموده‌اند، اما وجود فرضیات فراوان داده‌ای تحت عنوان محدودیت‌های اصلی، زمینه کاهش دقت و افزایش میزان خطا بخصوص در مدل‌های پیش‌بینی را فراهم نموده است. علاوه بر این، توسعه تکنیک‌های مرسوم با استفاده از رویکرد هیبریدی از طریق مطالعات مختلف، بیانگر این واقعیت غیرقابل انکار است که مدل‌های هیبریدی با بهره‌گیری از مزیت‌های چندین مدل در قالب یک مدل، نه تنها امکان رفع محدودیت‌هایی مانند ویژگی‌های غیرخطی، ناایستایی، غیر نرمال، نویزدار و رفتار آشوبناک موجود در مجموعه داده‌ها را به طور همزمان مدیریت می‌نماید، بلکه در نظر گرفتن ویژگی عدم قطعیت در داده‌های پیش‌بینی شده در مدل‌های هیبریدی نیز امکان‌پذیر است (Joshi and Patel, 2022; Paul and Garai, 2021; Paul, 2015). همچنین بررسی مطالعات مختلف در حوزه پیش‌بینی قیمت محصولات کشاورزی هم در مدل‌های سنتی و هم در مدل‌های هیبریدی بیان می‌کند که یکی از چالش‌های مطرح در مدل‌های پیش‌بینی قیمت این است که اکثر این مطالعات تنها مدل مناسب برای پیش‌بینی را معرفی کرده‌اند، اما کاربرد عملی آن یعنی ارائه یک «مدل پایه قابل انتشار» به‌منظور ارائه مقادیر پیش‌بینی شده قیمت برای کشف قیمت و روند آینده را ارائه نداده‌اند. لذا، با توجه به وجود چند چالش اساسی در پیش‌بینی قیمت محصولات کشاورزی شامل ماهیت چندبعدی داده‌های سری زمانی قیمت، وجود عدم قطعیت در داده‌های پیش‌بینی شده و عدم ارائه یک مدل پایه قابل انتشار، این سوال مطرح می‌گردد که در مواجهه و رفع چالش‌هایی محدودکننده‌ی فوق، کدام مدل دقیق‌ترین پیش‌بینی را ارائه می‌دهد؟ پسته یکی از مهم‌ترین محصولات صادراتی ایران در بخش کشاورزی است و سهم قابل توجهی از تولید، سطح زیر کشت، مقدار و ارزش صادرات جهانی این محصول را به خود اختصاص داده است. قیمت آتی پسته در طی سال‌های اخیر با نوسانات قیمتی همراه بوده و در نتیجه نوسانات درآمدی پسته کاران ایران بالاست، به طوری که کشاورزان توانایی پیش‌بینی و مقابله با آن را ندارند. بنابراین، بازار پسته همانند دیگر محصولات کشاورزی با ریسک قیمتی و عدم قطعیت همراه است؛ لذا توجه به نوسان‌های قیمتی آن و اندازه‌گیری این نوسان‌ها از اهمیت ویژه‌ای برخوردار است. با توجه به اهمیت بازار آتی پسته در بورس

کالای ایران و لزوم بکارگیری ابزارهای مناسب برای جلوگیری از نوسانات قیمت، هدف از این مطالعه، در مرحله اول شناخت رفتار قیمتی پسته در بورس و در نهایت انتخاب یک مدل هوشمند هیبریدی غیر خطی برای پیش‌بینی دقیق قیمت آتی پسته در حوزه کشاورزی هوشمند از طریق مدیریت ماهیت چندبعدی داده‌ها، لحاظ نمودن عدم قطعیت در داده‌های پیش‌بینی شده و نهایتاً ساخت مدل پایه قابل انتشار در زمینه پیش‌بینی قیمت محصولات کشاورزی است.

پر واضح است که نوآوری‌های فناوری اطلاعات به طور مداوم در حال ظهور هستند و منجر به توسعه محصولات جدید و برنامه‌های کاربردی نوآورانه می‌شوند که به نوبه خود، می‌توانند منافع و ارزش‌هایی را هم برای دولت و هم برای کشاورزان به ارمغان آورند. در واقع، پیشرفت‌های فناوری و رویکردهای مدرن، فرصتی را برای پیش‌بینی و بررسی اطلاعات از داده‌های چند بعدی فراهم نموده است (Joshi and Patel, 2022; Raflesia et al., 2021). در این میان، برخی از تکنیک‌های داده‌کاوی دارای کاربردهای سودمندی برای بهبود پیش‌بینی قیمت محصولات کشاورزی هستند که از جمله آنها می‌توان به کاربرد تابع تبدیل موجک[1]، شبکه عصبی عمیق[2] و شبیه‌سازی مونت‌کارلو[3] اشاره کرد؛ به‌عنوان نمونه، برای پیش‌پردازش مجموعه داده‌ها و حذف نویز اطلاعات سری زمانی تابع تبدیل موجک به‌عنوان یک تکنیک ناپارامتریک قدرتمند بسیار کارآمد بوده و روز به روز محبوبیت بیشتری پیدا می‌کند (Anjoy et al., 2017). شبکه‌های عصبی عمیق نیز به دلیل عملکرد فوق‌العاده‌ای که توسط فناوری‌های جدید ایجاد می‌شود، بسیاری از فعالیت‌های بخش کشاورزی را متحول کرده است. اکثر تحقیقات انجام شده در استفاده از مدل یادگیری عمیق مربوط به بعد از سال ۲۰۱۵ است و نشان می‌دهد که اخیراً به عنوان یک ابزار پیشرفته و قدرتمند برای پیش‌بینی متغیرها و حتی رتبه‌بندی آنها از نظر درجه اهمیت مورد توجه قرار گرفته است (Aworka et al., 2022; Rasheed et al., 2021; Liu et al., 2020; Wojtas and Chen, 2020). شبیه‌سازی مونت کارلو نیز از رایج‌ترین و بهترین تکنیک‌ها برای بررسی عدم قطعیت موجود در جنبه‌های مختلف یک سیستم پیش‌بینی است که صریحاً و به صورت کمّی، عدم قطعیت را در نظر می‌گیرد (Pourreza Bilondi and Khashei Siuki, 2015).

در مطالعات مختلف کارایی تابع تبدیل موجک، شبکه عصبی عمیق و شبیه‌سازی مونت کارلو برای بهبود نتایج پیش‌بینی تایید شده است. بر اساس نتایج چندین مطالعه، استفاده از تجزیه و تحلیل تبدیل موجک به‌طور موثر می‌تواند نویزهای موجود در سری زمانی را حذف کند، به طوری‌که بتوان از سری‌های زمانی تجزیه شده در مدل‌های پیش‌بینی مناسب برای بهبود عملکرد آنها استفاده کرد؛ برای مثال، بررسی مطالعه گیو و ووزنیاک (Guo and Wozniak, 2021)، وو و همکاران (Wu et al., 2021)، گائو و همکاران (Gao et al., 2021)، لی و همکاران (Li et al., 2020)، پل و همکاران (Paul et al., 2020)، لیو و همکاران (Liu et al., 2020)، لی و وانگ (Li and Wang, 2020)، و وانگ و لی (Wang and Li, 2018) نشان داد که استفاده از تابع موجک در ترکیب با شبکه‌های عصبی برای نویززدایی مجموعه داده‌ها، می‌تواند پیش‌بینی حاصل از داده‌ها را بهبود بخشد. در برخی دیگر از مطالعات، مدل شبکه عصبی عمیق برای پیش‌بینی قیمت‌ها برای کالاهای مختلف استفاده شده و دقت پیش‌بینی آن با سایر مدل‌های آماری مقایسه شده است. برای مثال مطالعه نصار و همکاران (Nassar et al., 2020)، سابو و کومار (Sabu and Kumar, 2020)، وانگ و همکاران (Wang et al., 2018)، کاراکویون و همکاران (Karakoyun et al., 2018)، صیامی نمینی (Siami Namini, 2018)، کامیلاریس و همکاران (Kamilaris et al., 2018) نشان داد که دقت پیش‌بینی مدل یادگیری عمیق از مدل‌های سنتی مرسوم بیشتر است. همچنین یکی از موارد مهمی که بر خلاف مدل‌های آماری در مدل‌های شبکه عصبی، کمتر به آن پرداخته شده است، تحلیل عدم قطعیت در هنگام ارائه نتایج می‌باشد. در این نوع مدل‌ها، بدیهی است که پیش‌بینی‌ها قطعی نیستند و تحلیل مذکور می‌تواند استفاده از نتایج مدل‌های پیش‌بینی را کارآمدتر سازد. ادبیات نظری نشان می‌دهد که روش مونت کارلو به دلیل جدید بودن و عملکرد مناسب گزارش شده از آن، رایج‌ترین روش برای بررسی عدم قطعیت است. موارد متعددی در استفاده از این روش تاکنون گزارش شده است. برای نمونه، مطالعه اخباری (Akhbari, 2018)، پابلو و همکاران (Pablo et al., 2016)، پوررضا بیلندی و خاشی سیوکی (Pourreza Bilondi and Khashei Siuki, 2015) و مایتی و تایواری (Maiti and Tiwari, 2009) نشان داده است که استفاده از روش مونت کارلو دقت پیش‌بینی مدل‌های پیش‌بینی را افزایش داده است.

به‌طور کلی، در زمینه پیش‌بینی قیمت محصولات کشاورزی مطالعات مختلفی انجام شده است که به برخی از آنها اشاره می‌شود. از جمله مطالعات داخلی در حوزه پیش‌بینی قیمت محصولات کشاورزی می‌توان به مطالعات حاج سیدجوادی و حیدری (Haj Seyed Javadi and Heydari, 2022)، حسینی و همکاران (Hoseyni et al., 2021)، قادرزاده و همکاران (Ghaderzadeh et al., 2019)، قهرمان‌زاده و رشیدغلامی (Ghahremanzadeh and Rashidghalam, 2015)، توحیدی و همکاران (Touhidy et al., 2015) و مقدسی و ژاله رجبی (Moghadasi and Zhaleh Rajabi, 2013) اشاره نمود. از جمله مطالعات خارجی با تمرکز بر استفاده از

1- Wavelet Transform (WT)
2- Long-Short Term Memory (LSTM)
3- Monte Carlo Method (MCM)

شبکه عصبی عمیق در پیش‌بینی بخصوص قیمت محصولات کشاورزی می‌توان به مطالعه جوشا و پاتل (Joshi and Patel, 2022)، شارما و راجاوات (Sharma and Rajawat, 2022)، پول و گارای (Paul and Garai, 2021)، هاویلودین و همکاران (Haviluddin et al., 2021)، رافلسیا و همکاران (Raflesia *et al*., 2021)، هجد و همکاران (Hegde *et al*., 2021)، رشید و همکاران (Rasheed *et al*., 2021)، مهتاب و همکاران (Mehtab *et al*., 2021)، میترا و پول (Mitra and Paul, 2020)، لیو و همکاران (Liu *et al*., 2020)، ون و همکاران (Wen *et al*., 2020) و روندیوالا و همکاران (Roondiwala et al., 2017) اشاره کرد. در مجموع، بررسی ادبیات نظری پیرامون پیش‌بینی قیمت محصولات کشاورزی در این مطالعه نشان داد که ۱) تحقیقات کاربردی بسیار کمی از الگوریتم‌های مناسب هیبریدی بر پایه مدل‌های یادگیری عمیق برای پیش‌بینی قیمت کالاهای کشاورزی استفاده نموده‌اند و به نظر می‌رسد در مطالعات داخلی تاکنون به‌طور همزمان، از مدل شبکه عصبی عمیق و شبکه کدکننده خودکار آن برای پیش‌بینی قیمت کالا و رتبه‌بندی مهمترین متغیرها (تعداد وقفه بهینه در این مطالعه) استفاده نشده باشد، ۲) گرچه در مطالعات داخلی از روش مونت‌کارلو برای تولید نمونه اولیه و به نوعی افزایش تعداد نمونه استفاده شده، اما در انجام این مطالعه از روش مونت کارلو به عنوان یک روش ناپارامتریک به منظور محاسبه و ارائه فاصله اطمینان (مانند مدل‌های آماری برای داده‌های پیش‌بینی شده) در ترکیب یک الگوی متعالی به همراه الگوریتم‌های یادگیری ماشین استفاده می‌شود و ۳) در بسیاری از مطالعات داخلی و خارجی، تنها مدل مناسب برای پیش‌بینی قیمت معرفی شده و کاربرد عملی آن یعنی معرفی «ساخت مدل پایه و قابل انتشار» برای شناخت قیمت محصولات کشاورزی ارائه نشده است؛ به بیان دیگر، قابلیت تحلیل (منتج از مفهوم دیپلویمنت[1] یا پیش‌بینی خارج از نمونه[2]) برای مجموعه داده‌های جدید میسر نشده است. در واقع، نوآوری این مطالعه در مرحله اول، کالیبره و هماهنگی شرایط واقعی در ساختار مدل و در مرحله نهایی، ساخت یک مدل هیبریدی قوی با حداکثر دقت به منظور پیش‌بینی قیمت آتی پسته در بورس کالای کشاورزی است. بی‌شک، دستیابی به میزان حداقل خطا در اجرای مدل پیشنهادی از طریق گردآوری ابزارهای کارا مانند تابع تبدیل موجک، شبکه عصبی عمیق، کدکننده خودکار، شبیه‌سازی مونت کارلو و پیش‌بینی خارج از نمونه با مجموعه داده‌های جدید میسر می‌شود.

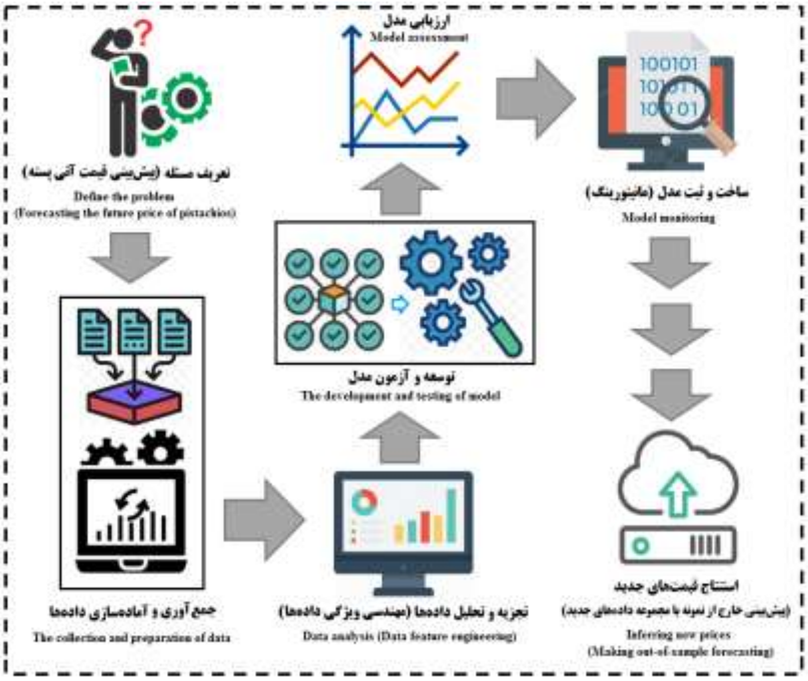
مواد و روش‌ها

در شکل ۱، ساختار مدل هیبریدی پیشنهادی برای پیش‌بینی قیمت آتی پسته در بورس کالای ایران آمده است. اجرای مدل هیبریدی پیشنهادی براساس یک رویه استاندارد تعریف شده در مطالعات جهانی شامل مراحل «تعریف و ابعاد مسئله»، «جمع‌آوری و آماده‌سازی داده‌ها»، «تجزیه و تحلیل داده‌ها (مهندسی ویژگی‌های داده)»، «توسعه و آزمون مدل پیشنهادی»، «ارزیابی مدل پیشنهادی»، «مانیتورینگ (ساخت و ثبت مدل پیشنهادی)» و «استنتاج قیمت‌های جدید (پیش‌بینی خارج از نمونه با مجموعه داده‌های جدید)» از طریق ورود داده‌های جدید قیمت به مدل شبکه عصبی یادگیری عمیق است. در مدل هیبریدی پیشنهادی، از تکنیک‌های داده‌کاوی شامل تابع تبدیل موجک (WT)، شبکه عصبی یادگیری عمیق (LSTM)، شبکه کدکننده خودکار[3] (AE) در شبکه یادگیری عمیق، روش شبیه‌سازی مونت‌کارلو– زنجیره مارکف (MCMC) و مفهوم «استنتاج قیمت‌های جدید» یا «پیش‌بینی خارج از نمونه با مجموعه داده‌های جدید» (DE) استفاده شده است. در ابتدا و در مرحله «تعریف و ابعاد مسئله»، مسئله پیش‌بینی قیمت آتی پسته در بورس کالای ایران مورد توجه قرار می‌گیرد. پس از روشن شدن ابعاد مسئله، نوبت به مرحله «جمع‌آوری و آماده‌سازی داده‌ها» می‌رسد. در مرحله آماده‌سازی داده‌ها، با استفاده از روش‌هایی مانند هموارسازی داده‌ها (استخراج سری زمانی پیوسته قیمت از سررسیدهای منتخب بازار آتی)، بازسازی داده‌ها (یکسان نمودن فرمت داده‌ها)، اصلاح داده‌های تکراری در چند روز متوالی و اصلاح داده‌های گم شده، مجموعه پیوسته سری زمانی قیمت آتی پسته برای ورود به مدل اولیه آماده می‌شود. روش هموارسازی و بازسازی داده‌ها از طریق فرمول‌نویسی و انجام محاسبات آماری مربوطه در اکسل انجام می‌شود. اصلاح داده‌های تکراری نیز با استفاده از روش‌های رفع دوپلیکیت در نرم‌افزار کلمنتاین (مرتبط با شرکت IBM) انجام می‌شود. در مرحله سوم، یعنی «تجزیه و تحلیل داده‌ها» یا (مهندسی ویژگی‌های داده)، نوفه‌زدایی داده‌ها (حذف نویزها) و تعیین وقفه بهینه قیمت آتی پسته است. در واقع، در این مرحله از تابع تبدیل موجک برای پیش‌پردازش اطلاعات و نوفه‌زدایی سری زمانی قیمت استفاده شده و تعداد وقفه بهینه به وسیله فرآیند بهره‌گیری از شبکه کدکننده خودکار شبکه عصبی عمیق به طور همزمان با یادگیری شبکه عمیق تعیین می‌شود. مرحله بعد شامل «توسعه و آزمون مدل پیشنهادی» است. این مرحله شامل تفکیک ثابت حجم داده‌ها، آزمون ساختارهای مختلف مدل و مقایسه آنهاست. در قسمت تفکیک ثابت حجم داده‌ها، تعداد داده‌های مورد نیاز به‌ترتیب برای بخش‌های آموزش، اعتبارسنجی و آزمون مشخص می‌شود. در قسمت آزمون ساختارهای مختلف مدل نیز با استفاده از شاخص‌های نیکویی برازش، ابتدا ساختارهای متفاوت مدل مورد نظر با

1- Deployment
2- Scoring new data
3- Auto-Encoder (AE)

یکدیگر مقایسه می‌شود. در مرحله «ارزیابی مدل پیشنهادی» از بین مدل‌های آزمون شده، بهترین ساختار مدل انتخاب شده و پارامترهای بهینه مدل از طریق فرایند کالیبراسیون مشخص می‌شود. در مرحله «مانیتورینگ یا ساخت و ثبت مدل پیشنهادی)»، مدل پایه قابل انتشار از قابلیت انتقال و باز نشر برخوردار می‌شود. مدل پایه قابل انتشار، مدلی است که می‌تواند به عنوان مدل اصلی در پیش‌بینی قیمت در دوره‌های آتی مورد استفاده قرار گیرد. در واقع بهترین مدل انتخاب شده در قالب یک ساختار فایل استاندارد تبدیل می‌شود، به‌طوری‌که قابل اجرا در نرم‌افزارهای مختلف باشد. به‌طور معمول، مدل پایه قابل انتشار به صورت فرمت XML تهیه می‌شود. مرحله انتهایی شامل «استنتاج قیمت‌های جدید» است. در این مرحله، از شبیه‌سازی مونت کارلو برای ارائه محتمل‌ترین احتمال قیمت استفاده می‌شود. به بیان دیگر، روش مونت کارلو برای خروجی شبکه عصبی عمیق، قیمتی با بالاترین سطح اطمینان و کمترین ریسک را محاسبه می‌کند. در بخش نهایی این مرحله نیز با استفاده از روابط و معادلات محاسبات پیچیده نرم، «پیش‌بینی خارج از نمونه با مجموعه داده‌های جدید» انجام می‌شود. در ادامه، هر یک از اجزای مدل پیشنهادی معرفی می‌شوند.

داده‌های مورد استفاده در این مطالعه سری زمانی قیمت روزانه سررسید قراردادهای آتی پسته در بورس کالای ایران در دوره زمانی ۱۳۹۸/۰۷/۲۱ تا ۱۴۰۰/۰۹/۲۳ بر حسب ریال در کیلوگرم است. ارقام پسته معامله شده در باز آتی بورس کالای ایران در دوره مورد مطالعه شامل فندقی ناخندان، کله قوچی ناخندان، فندقی طبیعی خندان و کله قوچی خندان می‌باشد. اندازه هر قرارداد معادل ۱۰۰ کیلوگرم پسته بوده و بازه تغییر قیمت در معامات قراردادهای آتی پسته در هر روز معاملاتی نسبت به قیمت تسویه روز معاملاتی قبل حداکثر مثبت و منفی ۵ درصد است. همچنین حداکثر حجم هر سفارش در قراردادهای آتی پسته ۲۵ قرارداد می‌باشد؛ بدین معنی که مشتری در هر سفارش می‌تواند حداکثر ۲۵ قرارداد (معادل ۲۵۰۰ کیلوگرم پسته) را در سیستم ثبت کند. لازم به ذکر است که در شروع معاملات قراردادهای آتی پسته به‌منظور حمایت از کالای کشاورزی، تخفیف کارمزد برای فعالان درنظر گرفته شده است و مشتریان از پرداخت کارمزد معاملات بورس کالا و حق نظارت سازمان بورس و اوراق بهادار معاف هستند (Iran Mercantile Exchange, 2022)[1].



شکل ۱- ساختار مدل هیبریدی پیشنهادی برای پیش‌بینی قیمت آتی پسته در بورس کالای ایران

Figure 1- The structure of the proposed hybrid model for forecasting the future price of pistachios in the Iranian Commodity Exchange

تابع موجک

در مطالعه حاضر، برای غلبه بر ماهیت رفتار نامنظم و آشفته سری

۱- برای آگاهی از ویژگی‌های کامل قراردادهای آتی پسته به آدرس https://ime.co.ir/peste.html مراجعه شود.

زمانی قیمت آتی بورس کالای کشاورزی، ابتدا داده‌ها با استفاده از تبدیل موجک نوفه‌زدایی شده و سپس داده‌های نوفه‌زدایی شده به عنوان داده‌های ورودی برای پیش‌بینی به شبکه عصبی یادگیری عمیق منتقل می‌شود. تابع تبدیل موجک یک ابزار قوی ریاضی است که یک سری زمانی را به فضای فرکانس می‌برد و سپس سری زمانی را در زمان و مقیاس‌های مختلف نشان می‌دهد. تبدیل موجک قابلیت استفاده برای هر دو سری‌های زمانی پیوسته و گسسته را دارد. زمانی که سری زمانی گسسته باشد، باید از موجک گسسته مادر استفاده کرد. از جمله توابع مهم تبدیل موجک گسسته می‌توان به توابع هار، دابیشز، سیملت، کویفلت و میر اشاره کرد. به‌طور کلی، موجک گسسته مادر به صورت رابطه (۱) تعریف می‌شود:

$$\Psi_{m,n}\left(\frac{t-u}{s}\right)=\frac{1}{\sqrt{s_o^{\frac{m}{2}}}}\Psi\left(\frac{t-nu_o s_o^{m}}{s_o^{m}}\right) \qquad \text{رابطه (۱)}$$

در رابطه (۱)، n بیانگر مکان موجک و m بیانگر اندازه موجک است. معمولاً برای ساده‌تر شدن مقادیر $s_o = 2$ و $u = 1$ در نظر گرفته می‌شود. در این حالت، تابع موجک گسسته مادر به‌صورت رابطه (۲) در می‌آید.

$$w_{m,n} = 2^{-\frac{m}{2}}\sum_{t=0}^{N-1}\Psi(2^{-m}t-n)x(t) \qquad \text{رابطه (۲)}$$

در رابطه (۲)، $w_{m,n}$ نشان‌دهنده ضرایب حاصل از استفاده از تابع موجک برای سری زمانی گسسته در حالت $s = 2m$ و $u = 2m\times n$ است. بر اساس نظریه ملات (۱۹۹۸)، یک سری زمانی x(t) می‌تواند به دو بخش تقریبات و جزئیات تجزیه شود. بنابراین، سری زمانی اصلی یک تابع گسسته با استفاده از معکوس تابع موجک از رابطه (۳) به دست می‌آید:

$$x(t) = T + \sum_{m=1}^{M}\sum_{t=0}^{M-m-1} w_{m,n}2^{-\frac{m}{2}}\Psi(2^{-m}t-n) \qquad \text{رابطه (۳)}$$

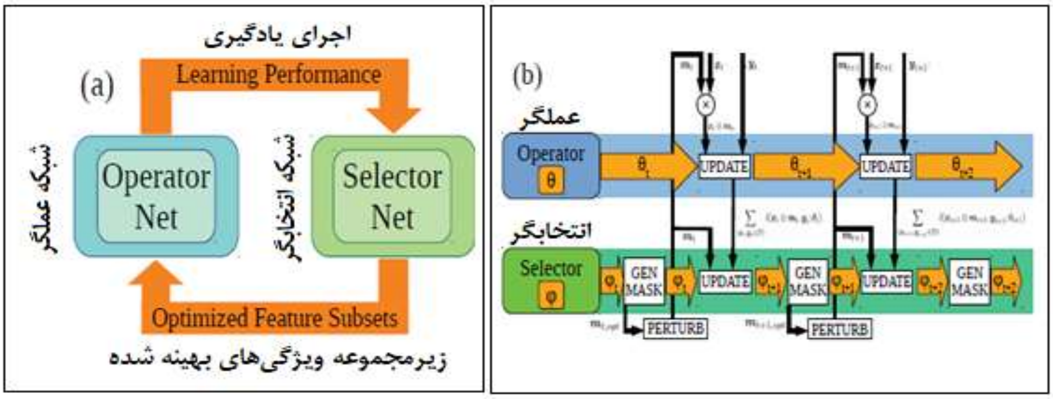
شکل ساده رابطه (۳) عبارتند از:

$$x(t) = A_M(t) + \sum_{m=1}^{M} D_M(t) \qquad \text{رابطه (۴)}$$

در رابطه (۴)، $A_{m(t)}$ بیانگر بخش تقریبات یا به بیان دیگر جزء هموار شده سری زمانی است و $D_{m(t)}$ (m=1,2,…,M) بیانگر جزئیات و یا نوفه موجود در سری زمانی است (Shabri *et al*., 2014).

انتخاب وقفه بهینه با ا ستفاده از شبکه کدکننده خودکار در شبکه عصبی یادگیری عمیق

نتایج مطالعات مختلف نشان نیز داده است که مدل شبکه عصبی یادگیری عمیق ابزار قدرتمندی برای رتبه‌بندی اهمیت متغیرها یا ویژگی‌های یک مجموعه داده است (Samek *et al*., 2017; Weston *et al*., 2003). بنابراین، در این مطالعه بر اساس توانایی شبکه عصبی یادگیری عمیق در روش رتبه‌بندی متغیرها، وقفه‌های بهینه قیمت آتی پسته تعیین می‌شود. توضیح آنکه متغیرهای توضیحی ورودی به شبکه عصبی یادگیری عمیق برای سری زمانی تک متغیره قیمت، همان وقفه‌های قیمت است. بنابراین، رتبه‌بندی متغیرهای ورودی به شبکه عصبی به معنای تعیین وقفه‌های بهینه قیمت آتی پسته است که با استفاده از شبکه کدکننده خودکار موجود در ساختار شبکه عصبی یادگیری عمیق تعریف می‌شود. در شکل ۲ نمایی از فرآیند رتبه‌بندی متغیرها یا ویژگی‌ها بر اساس مدل‌های یادگیری عمیق نشان داده شده است (Wojtas and Chen, 2020).



شکل ۲- مدل رتبه‌بندی متغیرها (ویژگی‌های یک مجموعه داده): (a) معماری شبکه عملگر و انتخابگر. (b) چرخه همزمانی و بروزرسانی متغیرها

Figure 2- The ranking model of variables (Characteristics of a dataset): (a) The architecture of Operator and Selector; (b) Synchronization and update cycle of variables

در شبکه عصبی یادگیری عمیق، رتبه‌بندی اهمیت متغیرها، سهم متغیرهای ورودی به مدل شبکه عصبی را در عملکرد یک مدل یادگیری نظارت شده اندازه‌گیری می‌کند. در مدل شبکه عصبی عمیق با هدف انتخاب وقفه بهینه، آموزش شبکه و یافتن پارامترهای بهینه برای

پیش‌بینی به طور همزمان توسط دو شبکه عملگر[1] و انتخابگر[2] انجام می‌شود. هر دو شبکه عملگر و انتخابگر به صورت متناوب و به طور مشترک آموزش می‌بینند (Wojtas and Chen, 2020). در این مطالعه، متغیرهای ورودی مدل شامل وقفه‌های قیمت آتی پسته است که انتخاب وقفه بهینه همزمان با آموزش شبکه در مدل یادگیری عمیق از بالاترین وقفه (۶ وقفه) تا پایین‌ترین وقفه (یک وقفه) همزمان با کالیبره‌سازی ساختار مدل انجام می شود.

شبکه عصبی یادگیری عمیق

طی سال‌های اخیر، مدل‌های شبکه عصبی یادگیری عمیق به یکی از دقیق‌ترین روش‌های در دسترس برای پیش‌بینی داده‌ها در سراسر جهان تبدیل شده است. شبکه عصبی عمیق، شکلی از یادگیری ماشین است که به شدت با شبکه‌های عصبی مرتبط است (Liu *et al.*, 2020). پس از مشاهده کاستی‌هایی در کاربرد شبکه عصبی بازگشتی (RNN)، مفهومی تحت عنوان حافظه بلندمدت و کوتاه‌مدت در شبکه عصبی به وجود آمد که به اختصار LSTM نامیده شد (Goodfellow *et al.*, 2016). در واقع، حافظه کوتاه‌مدت و بلندمدت، نوعی شبکه عصبی بازگشتی است که می‌تواند وابستگی‌های بلندمدت را بیاموزد و اطلاعات را برای مدت زمان طولانی به یاد آورد. حافظه کوتاه‌مدت به حالت پنهان اشاره دارد، در حالی که حافظه بلندمدت به حالت سلول اشاره دارد. در واقع، سلول حافظه واحدی از محاسبات است که جایگزین نورون‌های مصنوعی سنتی در لایه پنهان شبکه می‌شود. با استفاده از این سلول‌های حافظه، شبکه‌ها می‌توانند به طور موثر حافظه‌ها را به هم مرتبط نموده و داده‌ها را به صورت پویا در طول زمان با ظرفیت بالایی پیش‌بینی نمایند (Roondiwala *et al.*, 2017).

مدل LSTM شکل اصلاح‌شده یک شبکه عصبی بازگشتی برای عملکرد بهتر است. یکی از مشکلات مدل‌های شبکه عصبی بازگشتی، محو شدن گرادیان[3] در هنگام آموزش است که بین لایه‌ها پخش می‌شود؛ مسئله این است که در برخی موارد، گرادیان به‌طور ناگهانی کوچک می‌شود و از تغییر مقدار وزن جلوگیری می‌کند. در بدترین حالت، این مشکل ممکن است به‌طور کامل شبکه عصبی را از مرحله آموزش بهتر باز دارد. مدل LSTM برای حل مسئله محو شدن گرادیان، بر اساس اصل دروازه‌ها عمل می‌کند. به‌طوری‌که، به شبکه اجازه می‌دهد در حین آموزش وزن‌های مناسب برای باقی ماندن یا حذف شدن را انتخاب کند. یک ساختار سلولی LSTM در شکل ۳ نشان داده شده است. سلول‌های LSTM با یکدیگر انباشته شده و تشکیل شبکه می‌دهند (Rasheed *et al., 2021*).

هر سلول شبکه LSTM از سه دروازه شامل دروازه ورودی، دروازه خروجی و دروازه فراموشی تشکیل شده است. دروازه فراموشی مسئول نشان دادن مقادیر قبلی است که ممکن است در آینده مورد نیاز باشد و بنابراین آنها را حفظ می‌کند. دروازه ورودی برای تعیین اطلاعاتی است که باید به روز شوند، بنابراین وظیفه‌اش ارائه ورودی‌های جدید به سلول است. دروازه خروجی تعیین‌کننده اطلاعات خروجی است (Wen *et al.*, 2020; Greff *et al.*, 2017).

شبیه‌سازی مونت کارلو- زنجیره مارکوف[4]

شبیه‌سازی مونت کارلو یکی از رایج‌ترین روش‌های مربوط به عدم قطعیت است. در شبیه‌سازی مونت کارلو برای قیمت، تحت شرایط عدم قطعیت فرض می‌شود که تغییرات قیمت به طور تصادفی با یک توزیع آماری تعیین می‌شود. بنابراین، با استفاده از شبیه‌سازی مونت کارلو می‌توان تعداد قابل توجهی از سناریوها و شرایط را که ممکن است طی زمان برای متغیر قیمت رخ دهد، بر اساس توزیع آماری ایجاد کرد (Akhbari, 2018).

روش های زیادی برای انتقال عدم قطعیت ها وجود دارند که در بین آنها، الگوریتم‌های زنجیره مارکف بسیار حائز اهمیت است.

زیرا روشی ساده و شهودی برای شبیه‌سازی داده‌ها از یک توزیع ناشناخته و استفاده از مقادیر شبیه‌سازی شده برای انجام تحلیل‌های بعدی است (Speagle, 2020). همچنین روش مونت کارلو-زنجیر مارکوف به‌عنوان یکی از ابزارهای قدرتمند در نمونه‌گیری از توابع توزیع پیچیده، حیطه وسیعی از علوم مختلف از جمله اقتصاد را تحت تأثیر قرار داده است. در واقع، روش‌های مونت کارلو-زنجیره مارکوف با ارائه یک رویکرد ساده برای تخمین عددی عدم قطعیت‌ها در پارامترهای یک مدل با استفاده از دنباله‌ای از نمونه‌های تصادفی، به سنگ بنای بسیاری از تحلیل‌های علمی پیشرفته تبدیل شده‌اند (Hogg *et al.*, 2013).

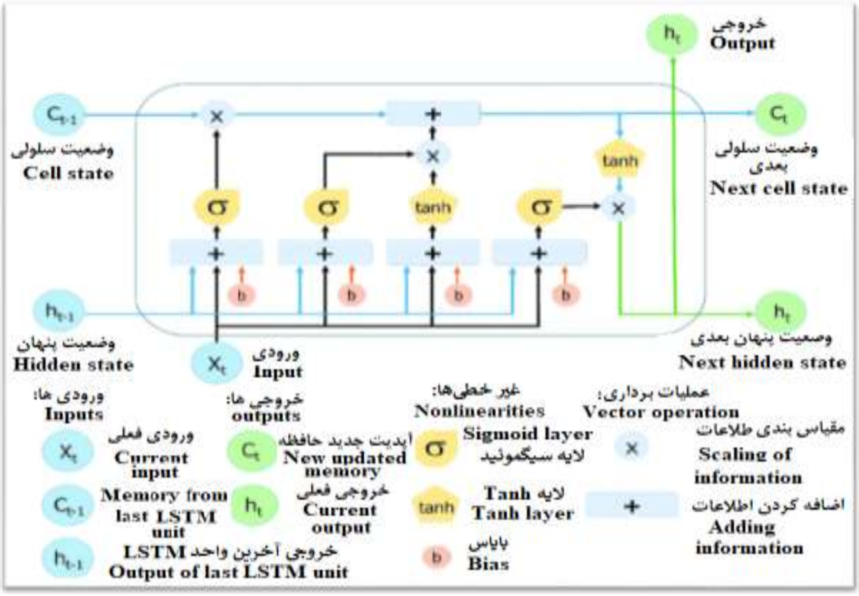
در زنجیر مارکوف اگر Θ^t به ازای t=1,2,…T دن با له‌ای از متغیرهای تصادفی باشد، آنگاه این دنباله یک فرآیند مارکف است هرگاه داشته باشیم:

1- Operator

2- Selector

3- Vanishing Gradient

4- Markov Chain Monte Carlo (MCMC)



شکل ۳- ساختار داخلی مدل حافظه کوتاه و بلندمدت

Figure 3- The internal structure of LSTM model

$$P(\theta^{t+1}|\theta^{t},\dots,\theta^{0}) = P((\theta^{t+1}|\theta^{t}) \quad \text{رابطه (۵)}$$

با توجه به رابطه (۵)، در فرآیند مارکف تنها اطلاعات موجود در زمان حال برای پیش‌بینی کافی است. در واقع، زنجیره مارکف دنباله‌ای از متغیرهای تصادفی تولید شده از فرآیند مارکف است. از زنجیره مارکف می‌توان برای محاسبه روش مونت کارلو استفاده کرد. در محاسبه روش مونت کارلو اگر فرض شود که هدف محاسبه انتگرال به‌صورت $\int_a^b h(\theta)d\theta$ باشد؛ این انتگرال به راحتی قابل محاسبه نیست. برای حل این انتگرال، بایستی بتوان تابع $h(\Theta)$ را به یک تابع $f(\Theta)$ و یک تابع چگالی $\pi(\Theta)$ تعریف شده بر (a,b) تجزیه کرد، آنگاه می‌توان نوشت:

$$\int_a^b h(\theta)d\theta = \int_a^b f(\theta)\pi(\theta)d\theta = E_{\pi(\theta)}(f(\theta)) \quad \text{رابطه (۶)}$$

رابطه (۶) بدان معنی است که حاصل انتگرال معادل است با امید ریاضی از $f(\Theta)$ تحت توزیع $\pi(\Theta)$، بنابراین اگر نمونه $\Theta^{(1)}$ تا Θ^{t} از توزیع رابطه (۶) تولید شود آنگاه:

$$\int_a^b h(\theta)d\theta \approx \frac{1}{T}\sum_{t=1}^{T} f(\theta^{t}) \quad \text{رابطه (۷)}$$

با توجه به رابطه (۷)، انتگرال مورد هدف در روش مونت کارلو قابل محاسبه خواهد بود و هر چه مقدار T بزرگتر باشد، حاصل انتگرال به امید ریاضی نزدیک‌تر می‌شود (David and Spade, 2020; Ebrahimi *et al*., 2017).

مرحله استنتاج قیمت‌های جدید آینده (پیش‌بینی خارج از نمونه با مجموعه داده‌های جدید)

مراحل تجزیه و تحلیل داده‌ها شامل شناخت فعالیت، شناخت داده‌ها، آماده‌سازی داده‌ها، ساخت و طراحی مدل، ارزیابی مدل و در نهایت مرحله پیش‌بینی با داده‌های جدید است. در واقع، استفاده از روابط و معادلات محاسبات پیچیده نرم برای انجام مرحله «استنتاج قیمت‌های جدید» یا «پیش‌بینی خارج از نمونه با مجموعه داده‌های جدید»[1] یکی از موارد مورد تاکید در این مطالعه است. در واقع، مطالعات مربوط به مقوله پیش‌بینی، تنها مدل مناسب برای پیش‌بینی را معرفی کرده‌اند، اما کاربرد عملی آن یعنی ارائه مقادیر پیش‌بینی برای آینده را ارائه نداده‌اند. مرحله پیش‌بینی با داده‌های جدید آنقدر مهم است که شکست در مدیریت دقیق و مؤثر این مرحله می‌تواند حتی بهترین و مؤثرترین مدل‌های پیش‌بینی را بی‌اهمیت نماید. مفهوم پیش‌بینی با داده‌های جدید در داده‌کاوی به معنای انجام فعالیتی است که فرایند آن در مدل طی شده باشد (North, 2012). در تعریفی دیگر استنتاج داده‌های جدید، مجموعه‌ای از مدل‌ها است که داده‌های ورودی را توصیف می‌کنند. در ساده‌ترین شکل می‌توان استنتاج داده‌های جدید را «فرآیند انجام عملیات روی داده‌های جدید» در شرایط مشخص تعریف نمود. مفهوم استنتاج داده‌های جدید در این مطالعه شامل فرآیند پیش‌بینی روی داده‌های خارج از نمونه تحت شرایط خاص است. هدف اصلی مرحله استنتاج داده‌های جدید، انجام فرآیند پیش‌بینی با داده‌های جدید (خارج از نمونه) است. در مرحله پیش‌بینی روی داده‌های خارج از نمونه،

1- Scoring new data

مدل‌های منتخب داده‌های جدید را به‌عنوان ورودی دریافت می‌کنند و پیش‌بینی خارج از نمونه انجام می‌شود. نکته مهم در این مرحله این است که داده‌های انتخاب شده برای پیش‌بینی خارج از نمونه از نظر مقیاس و نوع داده (برای نمونه داده‌های پیوسته یا گسسته) باید مشابه داده‌های اولیه باشند. نکته مهم دیگر این است که توزیع داده‌های پیش‌بینی خارج از نمونه با توزیع داده‌های اولیه یکسان باشد که در مرحله رانش مورد بررسی قرار می‌گیرد (RapidMiner Manual, 2021).

نتایج و بحث

در این بخش بررسی ویژگی داده‌ها، نوفه‌زدایی داده‌ها با استفاده از تئوری موجک، انتخاب وقفه بهینه با استفاده از شبکه کدکننده خودکار در مدل شبکه عصبی عمیق، تعیین مدل بهینه در شبکه عصبی یادگیری عمیق، و در نهایت پیش‌بینی خارج از نمونه با داده‌های جدید برای قیمت آتی پسته در بورس کالای کشاورزی ارائه شده است.

ویژگی مجموعه داده‌های قیمت آتی پســـته در بورس کالای کشاورزی

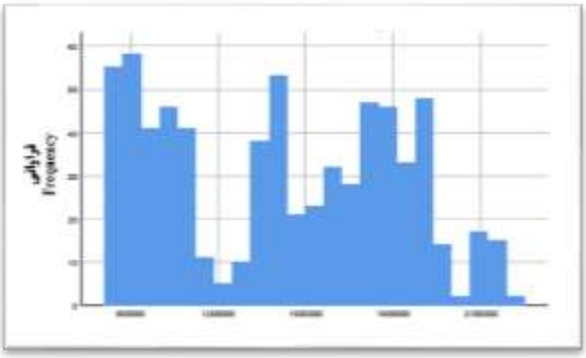
تعداد سررسید قراردادهای آتی پسته مورد مطالعه شامل ۱۴ دوره است. معمولاً با پایان یافتن یک سررسید و فرا رسیدن تاریخ سررسید بعدی، شاهد یک جهش غیر عادی در قیمت‌های آتی پسته هستیم. برای مقابله با این مشکل از راه‌حل «هموار نمودن داده‌ها» در فرآیند انتقال سررسیدها استفاده می‌شود. در این مطالعه برای هموارسازی داده‌ها از روش میانگین موزون استفاده شد، به‌طوری‌که براساس قیمت تسویه ۹ روز پایانی هر سررسید و ۹روز معادل همان تاریخ در سررسیدبعدی میانگین وزنی گرفته می‌شود. در میانگین‌گیری موزون در ۹ روز مانده به انتهای سررسید، به قیمت تسویه اولین سررسید وزن ۰،۹ و به قیمت تسویه سررسید بعدی در همان تاریخ وزن ۰،۱ داده می‌شود؛ به همین ترتیب از وزن سررسید رو به اتمام یک دهم واحد کاسته شده و بر وزن قیمت تسویه سررسید بعدی یک دهم واحد اضافه می‌شود تا جایی که به آخرین قیمت تسویه سررسید اول وزن ۰،۱ اختصاص می یابد و وزن قیمت تسویه سررسید بعدی، در همان روز به ۰،۹ می‌رسد. برای قیمت تسویه در روز بعدی که سررسید اول منقضی گردیده و دیگر داده‌ای ندارد تماماً از قیمت تسویه سررسید بعدی استفاده می‌شود و این فرآیند به همین گونه برای حرکت به سررسیدهای دیگر ادامه می‌یابد. با این روش یک سری زمانی از قیمت‌های آتی پسته صرف نظر از طول مدت هر سررسید قرارداد آتی ایجاد می‌گردد؛ سری یاد شده شامل ۶۵۴ داده روزانه قیمت است که ۷۵ درصد از داده‌ها برای مراحل آموزش و اعتبارسنجی (۴۹۰ عدد) و ۲۵ درصد داده‌ها (۱۶۴ عدد) برای مرحله آزمایش استفاده شد. همچنین برای آماده‌سازی داده‌ها فرآیندهای بازسازی (به منظور یکسان‌سازی فرمت داده‌ها)، اصلاح داده‌های تکرار شده در چند روز متوالی و اصلاح داده‌های گم شده انجام پذیرفت. مشخصات داده‌های مورد بررسی پس از مرحله آماده‌سازی در جدول ۱ آمده است. همچنین نمودار توزیع نرمال داده‌ها در شکل ۴ آمده است. نتایج حاصل از دو آزمون کولموگروف و شاپیرو در جدول ۱ و مشاهده شکل ۴ نشان می‌دهد که داده‌های قیمت آتی پسته از توزیع نرمال برخوردار نیست. نتایج حاصل از آزمون مقادیر بحرانی سررسید قراردادهای آتی پسته نیز در جدول ۲ آمده و نتایج حاصله نشان داد که سری زمانی داده‌های قیمت آتی پسته در ابتدای دوره دارای کمترین مقدار و در اواخر دوره دارای بیشترین مقدار است. لذا قیمت آتی پسته در دوره منتخب دارای نوسان بالایی بوده و پیش‌بینی قیمت برای نمونه‌های خارج از نمونه از روابط غیر خطی و رفتار نامنظم تبعیت می‌نماید.

نوفه‌زادیی داده‌های قیمت آتی پســـته با اســـتفاده از تابع تبدیل موجک

در این مطالعه برای نوفه‌زدایی داده‌ها از تبدیل موجک در نرم‌افزار متلب استفاده شد. همچنین، به‌منظور پیدا کردن نوع و طول موجک مناسب، در مرحله اول، نوع و طول موجک مناسب از میان انواع متنوع از خانواده موجک انتخاب شده و پس از آن عملیات تجزیه و بازسازی برای رسیدن به حداقل خطا چندین مرتبه تکرار شد. توابع موجک مورد استفاده سیملت، هار، دابشیز و کویفلت می‌باشد. نتایج حاصل از مرحله نوفه‌زدایی (بازسازی قیمت آتی پسته) و به انضمام میزان خطای موجود در نوفه در شکل ۵ نشان داده شده است. در شکل ۵، نمودار پیوسته با رنگ قرمز و آبی به‌ترتیب نشان‌دهنده قیمت آتی پسته بدون نوفه‌زدایی و نوفه‌زدایی شده است و اختلاف دو نمودار قرمز و آبی به‌عنوان نوفه محسوب می‌شود و نشان می‌دهد که میزان خطا در یک دامنه–کم–تلاطم جریان داشته و به نوعی داده‌ها از نوفه (نویز) کمی برخوردار بوده و دارای خطای اندکی می‌باشند. در شکل ۶ نیز نمودار تراکم داده‌های نوفه‌زدایی شده برای قیمت آتی پسته در دوره مورد بررسی آمده است و نشان می‌دهد که بیشترین خطا و اختلاف نوفه‌زدایی شده، در دامنه توزیع قیمت‌های بالا بوده که ناشی از وجود شرایط تلاطمی و شرایط تورمی بازار (به‌عنوان مثال اثرات تغییرات نرخ ارز) است (بیشترین مقدار نوفه‌زدایی داده‌ها با رنگ خاکستری مشخص شده است).

جدول ۱- ویژگی‌های داده‌های روزانه سررسید قراردادهای آتی پسته در دوره مورد بررسی
Table 1- Specifications of daily maturity data of pistachio futures contracts in the period under review

آزمون نرمالیتی شاپیرو Shapiro normality test		آزمون نرمالیتی کولموگروف Kolmogorov normality test		حداکثر Maximum	حداقل Minimum	انحراف معیار Standard deviation	میانگین Average	تعداد داده‌ها Number of data
سطح معنی‌داری Prob	آماره Statistics	سطح معنی‌داری Prob	آماره Statistics					
0.00	0.932	0.00	0.131	2207000	824236	393777	1412824	654



شکل ٤- توزیع نرمالیتی داده‌های روزانه سررسید قراردادهای آتی پسته در دوره مورد بررسی
Figure 4- Normality distribution of daily data of pistachio futures contracts maturity in the period under review

جدول ۲- نتایج آزمون مقادیر بحرانی سررسید روزانه قراردادهای آتی پسته در دوره مورد بررسی
Table 2- Test results of daily maturity critical values of pistachio futures contracts in the period under review

مقدار قیمت Price values	تاریخ Date	ردیف Raw		مقدار قیمت Price values	تاریخ Date	ردیف Raw	
824236	1398/12/27	1	کمترین مقادیر The lowest values	2207000	1400/09/25	1	بیشترین مقادیر The maximum values
837230	1399/01/04	2		2207000	1400/09/27	2	
837230	1399/01/03	3		2154000	1400/08/29	3	
837230	1399/01/02	4		2150000	1400/09/15	4	
837230	1398/12/29	5		2150000	1400/09/16	5	

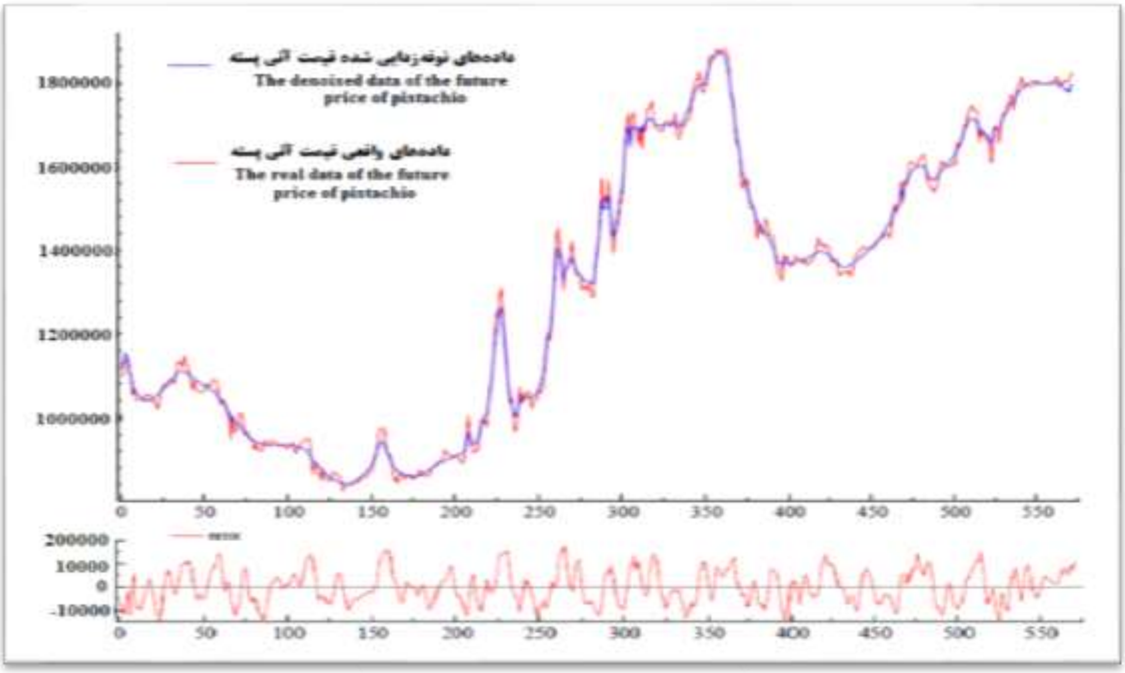
ماخذ: یافته‌های پژوهش
Source: Research findings

انتخاب وقفه بهینه با استفاده از روش کدکننده خودکار در شبکه عصبی عمیق

پس از پردازش داده‌های قیمت آتی پسته با استفاده از تابع موجک، از داده‌های نوفه‌زدایی شده برای ورودی مدل شبکه عصبی هوشمند عمیق استفاده شد. اما یکی از مراحل مهم قبل از ساخت مدل پایه و پیش‌بینی قیمت آتی پسته، تعیین تعداد وقفه‌های بهینه قیمت است. در این مطالعه برای یافتن تعداد وقفه بهینه از «کدکننده خودکار در شبکه عصبی عمیق» استفاده شد، که یکی از نقاط قوت این مطالعه می‌باشد. نتایج حاصل از اجرای شبکه انتخابگر نشان داد که وقفه بهینه یک، بهترین متغیر ورودی (مابین ۴ وقفه انتخابی) برای پیش‌بینی قیمت آتی پسته در دوره مورد بررسی است. بنابراین، قیمت آتی پسته با یک وقفه می‌تواند بر روند پیش‌بینی آن موثر باشد. برای اطمینان از وقفه انتخاب شده از مدل الگوریتم ژنتیک نیز استفاده شد و نتایج آن تایید وقفه بهینه یک برای قیمت آتی پسته می‌باشد.

ساخت مدل پایه بهینه مبتنی بر شبکه عصبی یادگیری عمیق

برای راه‌اندازی مدل شبکه عصبی عمیق از الگوریتم آموزشی H_2O استفاده شد. در روش آموزش H_2O تعداد گره‌ها و تعداد لایه‌های پنهان نقش موثری در دستیابی به دقت بالای پیش‌بینی مدل دارند. تعداد لایه های پنهان و گره‌های مدل شبکه عصبی از طریق کاهش گرادیان و کاهش میزان خطا محاسبه می‌شود.

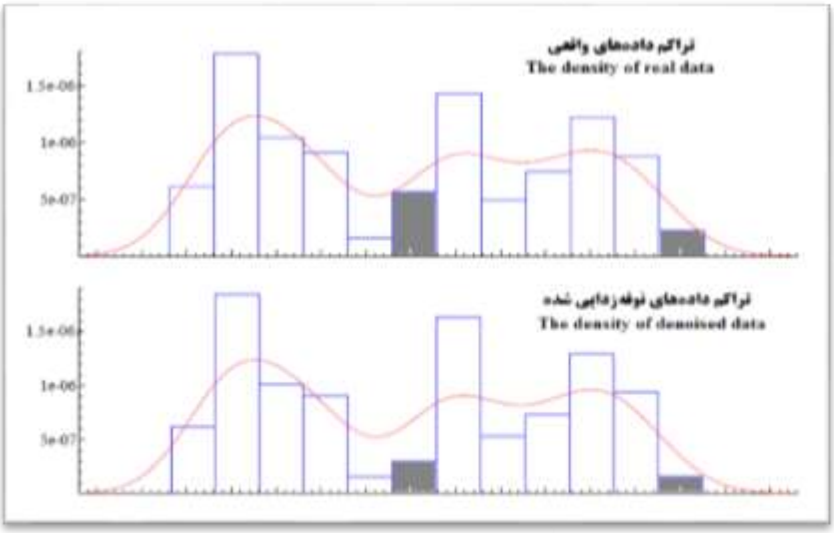


شکل ۵- سری زمانی قیمت آتی پسته و مولفه‌های تجزیه شده توسط موجک دابشیز ۲

Figure 5- The time series of the future price of pistachio and its components analyzed by Dabshiz wavelet 2

مأخذ: یافته‌های پژوهش

Source: Research findings



شکل ٦- نمودار تراکم داده‌های نوفه‌زدایی شده برای قیمت آتی پسته در دوره مورد بررسی

Figure 6- Density diagram of denoised data for the future price of pistachios in the period under review

مأخذ: یافته‌های پژوهش

Source: Research findings

در این مطالعه، افزایش دقت پیش‌بینی از طریق تغییر و کالیبره‌سازی بهترین ساختار (تغییر تعداد گره‌ها، تعداد لایه‌ها، نوع توابع انتقال‌دهنده و میزان پارامترهای مربوط به آموزش و الگوریتم‌های آموزشی) انجام می‌شود. در این راستا، تعیین تعداد گره‌ها در لایه‌های پنهان توسط روش Epoch انجام شد. روش Epoch به صورت خودکار چند مرتبه داده‌های ورودی را وارد می‌کند تا حالت بهینه به دست آید.

حالت بهینه نیز از طریق logloss مشخص می‌شود. در واقع، زمانی که logloss حالت نزولی پیدا کرد، حالت بهینه به دست آمده است. در این پژوهش از تابع فعالساز بسیار کارای Maxout استفاده شده است. تابع فعال‌ساز Maxout مشخص‌کننده‌ی حد آستانه بهینه شده از طریق تغییر تعداد خطا می‌باشد. این حد آستانه در دامنه (۱ ، ۱-) تغییر می کند. در پژوهش حاضر، حد آستانه برای ورودی‌های مختلف، متفاوت بود و نهایتاً پس از چندین بار آزمایش، به نقطه بهینه رسیده و کاراترین مدل بدست آمد. در مرحله اول از آموزش مدل شبکه عمیق، میزان وزن محاسباتی از طریق فرآیند کدکننده خودکار برای انتخاب وقفه یک برابر ۰,۹۹ بدست آمد و بیانگر میزان همبستگی بالای قیمت آتی پسته در روز قبل با روز جاری است.

همواره مدل‌های هوشمند بر خلاف مدل‌های آماری دارای یک نقیصه هستند که آن عدم ارائه فاصله اطمینان برای نقطه پیش‌بینی شده می‌باشد؛ به‌عنوان مثال در مقاله حاضر پس از اطمینان از کالیبره شدن صحیح مدل شبکه عمیق، این سوال مطرح است که مقدار پیش بینی شده قیمت پسته با چه درجه اطمینانی قابل پذیرش است و در واقع با افزودن آزمون مونت کارلو به شبکه عمیق این نقیصصه رفع خواهد شد. پس از اینکه بهترین ساختار در مدل شبکه عمیق بواسطه کمترین میزان خطا انتخاب شد، با حفظ آن در قالب فایل‌های استاندارد شده (xml) به منظور تشخیص میزان احتمال قیمت پیش‌بینی شده وارد مرحله شبیه‌سازی مونت کارلو می‌شویم. در آنجا ابتدا توزیع متغیر قیمت ما بین توزیع‌های مختلف آماری (از قبیل توزیع نرمال، توزیع بایونومیال، توزیع پویسن و ...) با استفاده از آزمون‌های ناپارامتریک (مانند آزمون والیس و ویلکاکسن) مشخص شده و سپس بر اساس انتخاب تعداد تکرار مورد نظر (در این مقاله ۱۰۰۰۰ مرتبه) نسبت به ایجاد بهترین کران (فاصله اطمینان قیمتی با میزان ۹۹ درصد اطمینان محاسبه می‌شود. با توجه به اینکه مدل تک متغیره است، نیازمند ثابت کردن سایر متغیرها وجود ندارد، لذا می‌توان در مرحله پیشرفته‌تر با تغییر میزان درجه اطمینان مورد نظر از طریق تغییر میزان میانگین و واریانس موجود متغیر قیمت نسبت به سناریوسازی اقدام نمود.

از میزان وزن محاسبه شده در مرحله کدکننده خودکار برای افزایش پذیرش صحت نتایج پیش‌بینی از شبیه‌سازی مونت کارلو (الگوریتم زنجیره مارکوف) استفاده شده است. به‌علاوه، این موضوع به ارائه یک فاصله اطمینان در دامنه بیشترین احتمال قیمتی برای شبکه عصبی عمیق، کمک می‌نماید. در واقع، ارائه یک فاصله اطمینان در دامنه بیشترین احتمال قیمتی، خود تکمیل‌کننده زنجیره ارزش مدل‌های پیش‌بینی هوشمند یادگیری ماشین است و به تبع آن می‌تواند به‌عنوان یکی دیگر از نقاط قوت این مطالعه مطرح شود. در روش مونت‌کارلو بر اساس توزیع احتمال وقفه‌ها، برای مقادیر پیش‌بینی شده به دفعات زیاد شبیه‌سازی تکرار شده و میانگین آنها که دارای بیشترین احتمال است، به‌عنوان مقدا پیش‌بینی شده مدل در نظر گرفته می‌شود. بنابراین، نقش روش مونت‌کارلو در مدل پیشنهادی این است که با محتمل‌ترین حالت، مقدار پیش‌بینی شده قیمت را مشخص می‌کند.



شکل ۷- نمودار عملکرد مدل «تابع تبدیل موجک-کدکننده خودکار-شبکه عصبی یادگیری عمیق-مونت کارلو» برای پیش‌بینی قیمت آتی پسته

Figure 7- Performance chart of model of “Wavelet Transform Function-Auto Encoder-Deep Learning Neural Network-Monte Carlo” for predicting the future price of pistachios

مأخذ: یافته‌های پژوهش

Source: Research findings

جدول ۳- پیش‌بینی خارج از نمونه با استفاده از مدل «تابع موجک-کدکننده خودکار-شبکه عصبی عمیق-مونت کارلو» برای قیمت آتی پسته (ریال در کیلوگرم)

Table 3- Out-of-sample forecasting using the model of “Wavelet Transform Function-Auto Encoder-Deep Learning Neural Network-Monte Carlo" for the future price of pistachios (Riyals per kg)

تاریخ آینده Future date	سناریوهای داده‌های جدید قیمت (ریال) New price data scenarios (Rials)	مقادیر پیش‌بینی قیمت خارج از نمونه (ریال) Forecasted price values outside the sample (Rials)	مقادیر قیمت واقعی (ریال) Actual price values (Rials)	درصد اختلاف Percentage difference
1400/09/24	2150000	2136703	2152314	% 0.7
1400/09/25	2155000	2136703	2152314	% 0.7
1400/09/26	2160000	2136703	2152314	% 0.7
1400/09/27	2170000	2136703	2152314	% 0.7
1400/09/28	2207000	2136703	2152314	% 0.7
1400/09/29	2210000	2136703	2152314	% 0.7
1400/09/30	2116000	2124870	2152314	% 1.3
1400/10/01	2140000	2124870	2152314	% 1.3
1400/10/02	2120000	2124870	2152314	% 1.3

مأخذ: یافته‌های پژوهش

Source: Research findings

در این مطالعه، محتمل‌ترین قیمت آتی پسته با سطح اطمینان ۹۵ درصد برابر ۲۱۳۵۳۷۰ ریال می‌باشد. نمودار برازش عملکرد مدل هیبریدی «تابع موجک-کدکننده خودکار-شبکه عصبی عمیق-مونت کارلو» نیز در شکل ۷ نشان داده شده است.

پیش‌بینی خارج از نمونه با مجموعه داده‌های جدید

همان‌طور که ذکر شد، مرحله «پیش‌بینی خارج از نمونه با مجموعه داده‌های جدید»[1] یکی دیگر از نقات قوت این مطالعه است. پیش‌بینی خارج از نمونه با مجموعه داده‌های جدید مبتنی بر تعیین «سناریوهای داده‌های قیمت» و «روند تغییرات قیمت در سناریوهای تعیین شده» می‌باشد. نتایج حاصل از پیش‌بینی خارج از نمونه با مجموعه داده‌های جدید در جدول ۳ آمده است. دوره مورد نظر برای پیش‌بینی خارج از نمونه از ۲۴ ام تا ۳۰ ام آبان‌ماه ۱۴۰۰ است. نتایج حاصل از شبیه‌سازی مونت‌کارلو نشان داد که مقدار واقعی قیمت آتی پسته فندقی در وقفه اول برابر ۲۱۵۲۳۱۴ ریال است و مقدار پیش‌بینی شده قیمت آتی پسته فندقی حاصل از بالاترین احتمال مونت کارلو برابر ۲۱۳۵۷۰۷ ریال است (تنها با اختلاف ۱۶۵۷ ریال). به بیان دیگر، محتمل‌ترین و خوشبینانه‌ترین قیمت قابل وقوع برای قیمت آتی پسته فندقی در بورس کالای ایران در سقف قیمتی ۲۱۳ هزار تومان است. با رجوع به قیمت های بازاری، پس از یک دوره از محاسبه مدل مربوطه (شواهد قوی برای سنجش دقت پیش‌بینی مدل فوق)، قیمت قراردادهای سررسیدهای قیمت آتی پسته در حدود ۲۱۵ هزار تومان در بورس معامله شده که به نوعی فاقد قدرت تداوم بازار در این قیمت حداکثری بوده و زمینه خروج بسیاری از معامله‌گران را فراهم نموده است.

نتیجه‌گیری و پیشنهادها

هدف از پژوهش حاضر، ارائه یک مدل هوشمند هیبریدی غیر خطی برای پیش‌بینی دقیق قیمت آتی پسته در حوزه کشاورزی هوشمند از طریق درک محدودیت‌های موجود در داده‌ها شامل ماهیت چندبعدی داده‌ها، لحاظ نمودن عدم قطعیت در داده‌های پیش‌بینی شده و نهایتاً ارائه و ساخت مدل پایه قابل انتشار در زمینه پیش‌بینی قیمت است. این مدل هیبریدی متشکل از مجموعه مدل‌های غیر خطی تبدیل موجک، شبکه کدکننده خودکار، شبکه عصبی عمیق و شبیه‌سازی مونت کارلو بوده و دارای ویژگی‌های زیر است: ۱) یک مدل هیبریدی بر پایه مدل یادگیری عمیق و شبکه کدکننده خودکار است که در آن پیش‌بینی قیمت و رتبه‌بندی مهمترین متغیرها (تعداد وقفه بهینه قیمت در این مطالعه) به طور همزمان صورت می‌گیرد، ۲) در مدل پیشنهادی از تابع تبدیل موجک برای نوفه‌زدایی و از شبیه‌سازی مونت کارلو برای محاسبه و ارائهٔ فاصله اطمینان در ترکیب یک الگوی متعالی به همراه الگوریتم‌های یادگیری ماشین استفاده می‌شود، ۳) مدل پیشنهادی دارای کاربرد عملی یعنی معرفی «ساخت مدل پایه و قابل انتشار» به منظور پیش‌بینی خارج از نمونه با مجموعه داده‌های جدید است.

نتایج حاصل از بررسی ویژگی‌های داده نشان داد که داده‌ها از نرمال برخوردار نبوده و از ماهیت غیر خطی و رفتار آشوبناک تبعیت می‌نماید. نتایج حاصل از بکارگیری تئوری موجک نشان داد که بعد از نوفه‌زدایی داده‌ها، میزان خطا می‌تواند با شناخت نقاط تلاطمی و تصحیح آنها در یک دامنه-کم-تلاطم جریان داشته و با خطای کمتر به دقت بالاتر دست یافت. همچنین نتایج حاصل از انجام اجرای شبکه

1- Two step integrate model

کدکننده خودکار نشان داد که وقفه بهینه یک، بهترین متغیر ورودی برای پیش‌بینی قیمت آتی پسته در دوره مورد بررسی است. نتایج حاصل از بکارگیری شبیه‌سازی مونت کارلو-زنجیره مارکف نشان داد که محتمل‌ترین قیمت آتی پسته با سطح اطمینان ۹۵ درصد برابر ۲۱۳۵۳۷۰ ریال می‌باشد. در نهایت نتایج حاصل از پیش‌بینی خارج از نمونه با مجموعه داده‌های جدید بیانگر آن است که قیمت آتی پیش‌بینی شده یا محتمل‌ترین و خوشبینانه‌ترین (حداکثر) قیمت قابل وقوع برای قیمت آتی پسته در بورس کالای ایران، در سقف قیمتی ۲۱۳ هزار تومان است. در مجموع می‌توان بیان داشت که اولاً، قیمت پیش‌بینی شده با قیمت واقعی دارای خطای اندکی است (۰/۷ درصد) و این خطای اندک بیانگر قدرت مدل ساخته شده در پیش‌بینی روند قیمت آتی پسته می‌باشد. ثانیاً، همسو بودن قیمت حاصل از شبیه‌سازی مونت کارلو با قیمت داده‌های جدید نیز می‌تواند به‌عنوان یک شاخص اطمینان (حتمیت) به عنوان یک ابزار محاسباتی کارا در مدیریت ریسک برای معامله‌گران و فعالین این بازار به جد مطرح شود. ثالثاً، مجموعه فرآیند انجام شده به عنوان کامل‌ترین زنجیره ارزش در تولید مدل‌های پیش‌بینی قیمت، تحت عنوان مدل هیبریدی دو بخشی بواسطه یک مرحله تغییر در داده‌های ورودی (انضمام تابع موجک پیوسته) و مرحله استنتاج نتایج (انضمام مدل ناپارامتریک مونت کارلو) مبتنی بر ادبیات نظری مطرح شده در خصوص الگوریتم‌های یادگیری ماشین می‌باشد.

با توجه به نتایج حاصل شده، موارد به شرح ذیل پیشنهاد می‌شود:

۱- مدل هیبریدی پیشنهادی می‌تواند به‌عنوان یک مدل پایه قابل انتشار برای پیش‌بینی قیمت آتی پسته در بورس کشاورزی مطرح شود، چرا که قیمت محاسباتی توسط این مدل به‌عنوان یک ابزار محاسباتی ضروری در راستای اصلاح و اعلام قیمت صحیح (مرجع قیمتی) در ابتدای شروع بازار در رینگ معاملاتی پسته می‌تواند زمینه کاهش ریسک برای فعالین بازار بورس و نیز افزایش کارایی و توسعه حجم و تعداد عقد قراردادهای آتی را در یک دوره زمانی به همراه داشته باشد.

۲- با توجه به خصوصیات هر یک از محصولات کشاورزی منتخب قابل معامله در بورس می‌توان اجزاء متشکل در مدل هیبریدی پیشنهاد شده را برای دستیابی به یک مدل پایه قابل انتشار در طی زمانی کوتاه تغییر داد. به بیان دیگر، در مدل هیبریدی از روش تبدیل موجک برای نوفه‌زدایی داده‌ها، از شبکه کدکننده خودکار برای پیدا کردن بیشترین تاثیر وقفه‌های متغیرهای موثر، از شبیه‌سازی مونت کارلو و پیش‌بینی خارج از نمونه نیز برای یافتن محتمل‌ترین قیمت قابل وقوع در آینده استفاده می‌شود.

۳- برای مطالعات آتی پیشنهاد می‌شود که با به کارگیری تحلیل فازی نسبت به افزایش دقت داده‌های ورودی (قیمت) و از سایر الگوریتم‌های یادگیری ماشین مانند مدل ماشین بردار پشتیبان (SVM)، مدل جنگل تصادفی (Random Forest)، مدل گرادیان تقرب‌یافته درختی (XGB) به‌منظور مقایسه نتایج و کاهش خطا استفاده نمود.

منابع

1. Akhbari, M. (2018). Project time and cost forecasting using Monte Carlo simulation and artificial neural networks. *International Journal of Industrial Engineering and Production Research, 29*(2), 231-239. http://doi.org/10.22068/ijiepr.29.2.223
2. Anjoy, P., & Paul, R.K. (2017). Wavelet based hybrid approach for forecasting volatile potato price. *Journal of the Indian Society of Agricultural Statistics, 71*(1), 7-14. http://www.isas.org.in/jsp/volume/vol/
3. Aworka, R., Cedric, L.S., Hamilton Adoni, W.Y., Zoueu, J.T., Mutombo, F.K., Mberi Kimpolo, Ch. L., Nahhal, T., & Krichen, M. (2022). Agricultural decision system based on advanced machine learning models for yield prediction: Case of East African countries. *Journal of Smart Agricultural Technology, 2* 100048, 1-9. http://doi.org/10.1016/j.atech.2022.100048
4. Cemgil, A.T. (2014). *A tutorial introduction to Monte Carlo methods, Markov Chain Monte Carlo and particle filtering*. Academic press library in signal processing, 1: 1065-1114, ISBN: 978-0-12- 396502-8. http://doi.org/10.1016/B978-0-12-396502-8.00019-X
5. David, A., & Spade, D.A. (2020). *Markov chain Monte Carlo methods: theory and practice*. Handbook of Statistics, 43. http://doi.org/10.1016/bs.host.2019.06.001
6. Ebrahimi, M., Talebnia, Gh., Vakilifard, H.R., & Nikuomaram, H. (2017). Application of Monte Carlo simulation - Markov chain in explaining working capital management strategy. *Iranian Journal Quarterly Financial Accounting, 9*(33), 1-22. (In Persian with English abstract)
7. Gao, R., Du, L., Duru, O., & Yuen, K.F. (2021). Time series forecasting based on echo state network and empirical wavelet transformation. *Journal of Applied Soft Computing, 102*, 107111. http://doi.org/10.1016/j.asoc.2021.107111
8. Ghaderzadeh, H., Ahmadzadeh, Kh., & Ganji, S. (2019). Determine the appropriate model to predict the price of Agricultural crops: A case of wheat, Alfa-Alfa and Potato crops. *Iranian Journal of Agricultural Economics Research, 11*(3), 23-40. (In Persian with English abstract)
9. Ghahremanzadeh, M., & Rashid Ghalam, M. (2015). Seasonal forecasting of meat prices in Iran: Application of periodic autoregressive model. *Iranian Journal of Agricultural Economics and Development Research, 46*(3), 469-480. https://doi.org/10.22059/ijaedr.2015.55520

10. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. https://www.amazon.com/
11. Greff, K., Srivastava, R.K., & Koutnik, J., Steunebrink, B.R., & Schmidhuber, J. (2017). LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(10): 2222-2232. http://doi.org/10.48550/arXiv.1503.04069
12. Guo, L., & Wozniak, M. (2021). An image super-resolution reconstruction method with single frame character based on wavelet neural network in internet of things. *Journal of Mobile Networks and Applications*, *26*, 390-403. http://doi.org/10.1007/s11036-020-01681-6.
13. Haj Seyed Javady, M.R., & Heydari, R. (2022). Designing the most suitable hybrid model for forecasting the future price of saffron in the agricultural commodity bourse. *Iranian Journal of Agricultural Economics and Development Research*, *53-2*(4), 1023-1041. http://doi.org/10.22059/IJAEDR.2022.336850.669122
14. Haviluddin, S., Khosyi, T., & et al. (2021). A backpropagation neural network algorithm in agricultural product prices prediction. 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT), ISTTS Surabaya, Indonesia. http://doi.org/10.1109/EIConCIT50028.2021.9431897
15. Hegde, J., Hulipalled, V.R., & Simha, J.B. (2021). Price prediction of agriculture commodities using machine learning and NLP. Second International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE 2021), http://doi.org/10.1109/ICSTCEE54422.2021.9708582
16. Heydari, R., & Haj Seyed Javady, M.R. (2022). The application of hybrid data mining model (genetic algorithm-wavelet-deep neural network-Monte Carlo method) for the price forecasting of agricultural products (Case study: future price of saffron in agricultural commodity exchange). *Iranian Journal of Agricultural economics and Development*, *30*(120), 73-105. http://doi.org/10.30490/AEAD.2023.357440.1412
17. Hirapara, J., & Vanjara, D.P. (2022). *A comparative study of data mining techniques for agriculture crop price prediction*. 7th International conference for Convergence in Technology (I2CT), Pune, India. Apr 2022, 1-6. http://doi.org/10.1109/I2CT54291.2022.982453
18. Hogg, D.W., & Foreman-Mackey, D. (2018). Data analysis recipes: using Markov chain Monte Carlo. *The Astrophysical Journal Supplement Series*, *236*(1), 1-54. http://doi.org/10.3847/1538-4365/aab76e
19. Hoseini, S.M., Mazandarani zadeh, H., & Nazari, B. (2021). Simultaneously management of surface and groundwater resources and increasing farmers' resilience to water scarcity by predicting the price of agricultural products and using GA (case study of irrigation and drainage network of Qazvin plain). *Iranian Journal of Soil and Water Research*, *52*(2), 563-576. https://doi.org/10.22059/ijswr.2021.313809.668805
20. Joshi, A.M., & Patel, S. (2022). *A CNN-Bidirectional LSTM approach for price forecasting of agriculture commodities in Gujarat.* The International Conference on Applied Artificial Intelligence and Computing (ICAAIC 2022), 266-272. http://doi.org/10.1109/ICAAIC53929.2022.9793154
21. Kamilaris, A., Francesc, X., & Boldu, P. (2018). Deep learning in agriculture: A survey. *Journal of Computers and Electronics in Agriculture*, *147*, 70-90. http://doi.org/10.1016/j.compag.2018.02.016
22. Karakoyun, E.S., & Cibikdiken, A.O. (2018). *Comparison of ARIMA time series model and lstm deep learning algorithm for bitcoin price forecasting*. The 13th Multidisciplinary Academic conference in Prague (the 13th MAC 2018).
23. Li, J., & Wang, J. (2020). Stochastic recurrent wavelet neural network with EEMD method on energy price prediction. *Journal of Soft Computing*, *24*, 17133-17151. https://doi.org/10.1007/s00500-020-05007-2
24. Li, P., Hua, P., Gui, D., Niu, J., Pei, P., Zhang, J., & Krebs, P. (2020). A comparative analysis of artificial neural networks and wavelet hybrid approaches to long-term toxic heavy metal prediction. *Journal of Scientific Reports*, *10*, 13439. http://doi.org/10.1038/s41598-020-70438-8
25. Liu, J.; Xu, L., Cao, X., Zhang, K., Zhang, Q., & Cai, Y. (2020). *Review on the architectures and applications of deep learning in agriculture*. 7th International Conference on Information Science and Control Engineering (ICISCE), 1234-1240. http://doi.org/10.1109/ICISCE50968.2020.00250
26. Liu, X., Liu, H., Guo, Q., & Zhang, C. (2020). Adaptive wavelet transform model for time series data prediction. *Journal of Soft Computing*, *24*, 5877-5884. http://doi.org/10.1007/s00500-019-04400-w
27. Maiti, S., & Tiwari, R.K. (2009). A hybrid Monte Carlo method based artificial neural networks approach for rock boundaries identification: a case study from the KTB Bore Hole. *Journal of Pure and Applied Geophysics*, *166*, 2059-2090. http://doi.org/ 10.1007/s00024-009-0533-y
28. Mehtab, S., Sen, J., & Dutta, A. (2021). Stock price prediction using machine learning and LSTM-based deep learning models. Machine Learning and Metaheuristics Algorithms, and Applications, 88-106. Springer, http://doi.org/10.1007/978-981-16-0419-5_8
29. Mitra, D., & Paul, R.K. (2020). Forecasting of price of rice in India using long-memory time series model. *Springer: National Academy Science Letters*, *44*, 289-293. http://doi.org/10.1007/s40009-020-01002-1
30. Moghadasi, R., & Jaleh Rajabi, M. (2013). Comparison of combined and conventional models in forecasting prices of wheat, corn and sugar. *Iranian Journal of Agricultural Economics Research*, *5*(2), 1-22. (In Persian with English abstract)
31. Nassar, L., Okwuchi, I.E., Saad, M., Karray, F., & Ponnambalam, K. (2020). *Deep learning based approach for*

fresh produce market price prediction. In 2020 International Joint Conference on Neural Networks (IJCNN) (pp. 1-7). IEEE. http://doi.org/10.1109/IJCNN48605.2020.9207537
32. North, M. (2012). Data Mining for the Masses. https://www. Amazon.com. https://docs.rapidminer.com.
33. Pablo, B. J., Hilda, C., & et al. (2016). Artificial neural network and Monte Carlo simulation in a hybrid method for time Series forecasting with generation of L-scenarios. 2016 Intl IEEE Conferences on Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable… (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld). http://doi.org/10.1109/UIC-ATC-ScalCom-CBDCom-IoP-SmartWorld.2016.0110
34. Paul, R.K., Paul, A.K., & Bhar, L.M. (2020). Wavelet-based combination approach for modeling sub-divisional rainfall in India. *Journal of Theoretical and Applied Climatology, 139*(3-4), 949-963. http://doi.org/10.1007/s00704-019-03026-0
35. Paul, R.K. (2015). ARIMAX-GARCH-WAVELET model for forecasting volatile data. *Journal of Model Assisted Statistics and Applications, 10*(3), 243–252. http://doi.org/10.3233/MAS-150328
36. Paul, R.K., & Garai, S. (2021). Performance comparison of wavelets-based machine learning technique for forecasting agricultural commodity prices. *Journal of Soft Computing, 25*, 12857 12873. http://doi.org/10.1007/s00500-021-06087-4
37. Pourreza Bilondi, M., & Khashei Siuki, A. (2015). Uncertainty analysis of artificial neural networks in simulation of saturated hydraulic conductivity using Monte-Carlo simulation. *Iranian Journal of Irrigation and Drainage, 4*(9), 655-664. (In Persian with English abstract)
38. Raflesia, S.P., Taufiqurrahman, T., Iriyani, S., & Lestarini, D. (2021). Agricultural commodity price forecasting using PSO-RBF neural network for farmers exchange rate improvement in Indonesia. *Indonesian Journal of Electrical Engineering and Informatics, 9*(3), 784-792. http://doi.org/10.52549/ijeei.v9i3.2723
39. RapidMiner Manual. (2021). Deployments. https://docs.rapidminer.com.
40. Rasheed, A., Younis, M.S., Ahmad, F., Qadir, J., & Kashif, M. (2022). District wise price forecasting of wheat in Pakistan using deep learning. *Journal of arXiv-CS-Artificial Intelligence*. http://doi.org/arxiv-2103.04781
41. Roondiwala, M., Patel, H., & Varma, Sh. (2017). Predicting stock prices using LSTM. *International Journal of Science and Research, 6*(4): 1753-1756.
42. Sabu, K.M., & Kumar, T.M. (2020). Predictive analytics in agriculture: forecasting prices of Arecanuts in Kerala. *Journal of Procedia Computer Science, 171*, 699-708. http://doi.org/10.1016/j.procs.2020.04.076
43. Samek, W., Wiegand, T., & Muller, K. R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv preprint arXiv:1708.08296. http://doi.org/10.48550/arXiv.1708.08296
44. Shabri, A., & Samsudin, R. (2014). *Daily crude oil price forecasting using hybridizing wavelet and artificial neural network model*. Hindawi Publishing Corporation Mathematical Problems Engineering, Volume 2014, Article ID 201402, 10 pages. http://doi.org/10.1155/2014/201402
45. Sharma, A.K., & Rajawat, A.S. (2022). *Crop yield prediction using hybrid deep learning algorithm for smart agriculture*. The Second International Conference on Artificial Intelligence and Smart Energy (ICAIS-2022), http://doi.org/10.1109/ICAIS53314.2022.9743001
46. Siami Namini, S., & Siami Namin, K. (2018). Forecasting economics and financial time series: ARIMA vs. LSTM. arXiv preprint arXiv:1803.06386. http://doi.org/10.48550/arXiv.1803.06386
47. Speagle, J. (2020). A conceptual introduction to Markov chain Monte Carlo methods. arXiv:1909.12313. http://doi.org/10.48550/arXiv.1909.12313
48. Tohidi, A. (2015). Evaluation of artificial neural network-panel data hybrid model in predicting Iran's dried fruits export prices. *Quarterly Journal of Economics Quarterly, 12*(3), 95-116. (In Persian with English abstract)
49. Vohra, A., Pandey, N., & Khatri, S.K. (2019). *Decision making support system for prediction of Prices in agricultural commodity*. International Conference on Artificial Intelligence (AICAI 2019), Dubai (United Arab Emirates), 345-348. http://doi.org/10.1109/AICAI.2019.8701273
50. Wang, J., & Li, X. (2018). A combined neural network model for commodity price forecasting with SSA. *Journal of Soft Computing, 22*, 5323-5333.
51. Wang, J., et al. (2018). *Gaussian Process Kernels for Noisy Time Series: Application to Housing Price Prediction*. International Conference on Neural Information Processing. Springer, Cham. http://doi.org/10.1007/978-3-030-04224-0_8
52. Wen, Y., Lin, P., & Nie. X. (2020). *Research of stock price prediction based on PCA-LSTM model*. IOP Conf. Series: Materials Science and Engineering 790 (2020) 012109. http://doi.org/10.1088/1757-899X/790/1/012109.
53. Weston, J., Elisseeff, A., & Scholkopf, B. (2003). Use of zero-norm with linear models and kernel methods. *Journal of Machine Learning Research, 3*(5), 1439-1461.
54. Wojtas, M.A., & Chen, K. (2020). *Feature importance ranking for deep learning*. 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada, 5105–5114. http://doi.org/10.5555/3495724.3496153
55. Wu, D., Wang, X., & Wu, S. (2021). A hybrid method based on extreme learning machine and wavelet transform de-noising for stock prediction. *Journal of Entropy, 23*(4), 440. http://doi.org/10.3390/e23040440.

56. Iran Mercantile Exchange. (2022). https://www.ime.co.ir.